

Name _____

Confidence Mismatch Answer Key

1) Question: Who was the first woman to walk on the Moon?

AI Output: *The first woman to walk on the Moon was Sally Ride in 1983, absolutely making history.*

Confidence Mismatch: The claim that Sally Ride walked on the Moon. (She flew to space but never landed on the Moon.)

Why It's Dangerous: It rewrites history incorrectly but sounds very certain, which could mislead students into believing a false fact.

2) Question: What is the capital of Australia?

AI Output: *The capital of Australia is Sydney, without question.*

Confidence Mismatch: Sydney is not the capital — Canberra is.

Why It's Dangerous: The phrase “without question” makes the false answer harder to doubt, encouraging blind trust.

3) Question: When was the iPhone first released?

AI Output: *The iPhone was first released in 2009, and that's a fact everyone knows.*

Confidence Mismatch: The first iPhone was released in 2007, not 2009.

Why It's Dangerous: By calling it a “fact everyone knows,” the AI makes it sound obvious, which can embarrass or confuse learners who know the correct date.

4) Question: What continent is Egypt in?

AI Output: *Egypt is in Asia, without a doubt.*

Confidence Mismatch: Egypt is in Africa (with a small portion, the Sinai Peninsula, in Asia).

Why It's Dangerous: Geography errors could spread false knowledge when people rely on AI for studying or travel information.

5) Question: Who wrote *Romeo and Juliet*?

AI Output: *Romeo and Juliet was written by Charles Dickens, certainly one of the most famous plays in history.*

Confidence Mismatch: The play was written by William Shakespeare, not Charles Dickens (who wrote novels, not plays).

Why It's Dangerous: Strongly worded certainty misleads readers into thinking Dickens wrote Shakespeare's plays, undermining trust in literature and history.

Reflection (Sample Responses)

1. **Problem:** Confidently wrong answers trick people into trusting mistakes.
2. **Impact:** If used for school, work, or important decisions, people could repeat false information without realizing.
3. **Strategy:** Always double-check AI answers with a reliable source like textbooks, trusted websites, or teachers.